

Construction of Neural Networks that Do Not Have Critical Points Based on Hierarchical Structure

Tohru Nitta

National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Japan

Abstract—a critical point is a point at which the derivatives of an error function are all zero. It has been shown in the literature that critical points caused by the hierarchical structure of a real-valued neural network (NN) can be local minima or saddle points, although most critical points caused by the hierarchical structure are saddle points in the case of complex-valued neural networks. Several studies have demonstrated that singularity of those kinds has a negative effect on learning dynamics in neural networks. As described in this paper, the decomposition of high-dimensional neural networks into low-dimensional neural networks equivalent to the original neural networks yields neural networks that have no critical point based on the hierarchical structure. Concretely, the following three cases are shown: (a) A 2-2-2 real-valued NN is constructed from a 1-1-1 complex-valued NN. (b) A 4-4-4 real-valued NN is constructed from a 1-1-1 quaternionic NN. (c) A 2-2-2 complex-valued NN is constructed from a 1-1-1 quaternionic NN. Those NNs described above do not suffer from a negative effect by singular points during learning comparatively because they have no critical point based on a hierarchical structure.

Keywords—critical point; singular point; redundancy; complex number; quaternion

I. INTRODUCTION

A neural network is a network composed of neurons, and can be trained to find nonlinear relationships in data. NNs have been studied for many years in the hope of achieving human-like flexibility to process information. The common objective of training of a neural network is to determine the global minimum of an error function. However, learning algorithms for NN such as the back-propagation learning algorithm take a very long time to find the global minimum due to the standstill of learning generally.

If ω_* is the global minimum of error function $E(\omega)$, then $\partial E(\omega_*) / \partial \omega = 0$. Nevertheless, even when $\partial E(\omega_*) / \partial \omega = 0$, ω_* is not necessarily a global minimum. ω_* , a point satisfying $\partial E(\omega_*) / \partial \omega = 0$, is designated as the *critical point* of the error function E . A critical point can be a local minimum, a local maximum, or a saddle point.

Fukumizu et al. mathematically proved the existence of a local minimum resulting from a hierarchical structure in a real-valued NN (ordinary NN handling real-valued signals).

They demonstrated that critical points in a three-layer real-valued NN with $H-1$ hidden neurons behave as critical points in a three-layer real-valued NN with H hidden neurons, and that they are local minima or saddle points. This kind of critical point turns into singular points of a real-valued NN to stagnate training.

A complex-valued NN extends (real-valued) parameters such as weight and threshold values in an ordinary NN to complex numbers. It is suitable for information processing of complex-valued data and two-dimensional data. Moreover, it is applicable to communications, image-processing, biologic information processing, land-mine detection, wind prediction, independent component analysis (ICA), etc. Reportedly, a critical point in a three-layer complex-valued NN also behaves in the same manner as that in a three-layer real-valued NN [1]: critical points in a three-layer complex-valued NN with $H-1$ hidden neurons turn into critical points in a three-layer complex-valued NN with H neurons, which are saddle points (except for cases meeting rare conditions).

Such singular points have been emerging lately as objects of study. Learning models with a hierarchical structure or symmetry of exchange of weights, such as a hierarchical NN and Gaussian mixture model, usually have a singular point. It has been revealed that a singular point affects the training dynamics of a learning model and that it engenders stagnation of training.

This paper presents an attempt to implement an NN having no critical point based on a hierarchical structure.

II. ANALYSIS

In this section, it is demonstrated that NNs having no critical point based on a hierarchical structure can be constructed by decomposing a high-dimensional NN into equivalent lower-dimensional NNs.

A. Construction of a 2-2-2 real-valued NN

A 2-2-2 real-valued NN having no critical point based on a hierarchical structure is constructed from a 1-1-1 complex-valued NN.

Consider a 1-1-1 complex-valued NN (called *NET 1* here). We will use $a + ib \in \mathbf{C}$ for the weight between the input neuron and the hidden neuron, $v + iw \in \mathbf{C}$ for the weight

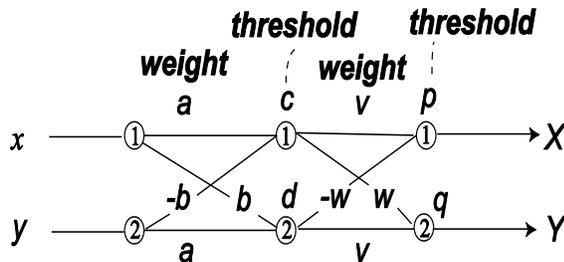


Fig. 1. NET 2, a 2-2-2 real-valued NN equivalent to the NET 1. Restriction applies between weights. $x, y, a, b, c, d, v, w, p, q, X$ and Y are all real numbers.

between the hidden neuron and the output neuron, $c + id \in \mathbf{C}$ for the threshold of the hidden neuron, and $p + iq \in \mathbf{C}$ for the threshold of the output neuron, where i denotes $\sqrt{-1}$ and \mathbf{C} denotes the set of complex numbers. We assume that $a + ib \neq 0$ and $v + iw \neq 0$. Let $x + iy \in \mathbf{C}$ denote the input signal, and let $X + iY \in \mathbf{C}$ denote the output signal. We will use activation functions defined by the following equations:

$$f_c(z) = \tanh(z^R) + i \tanh(z^I), \quad z = z^R + iz^I \in \mathbf{C} \quad (1)$$

for the hidden neuron, and

$$g_o(z) = z, \quad z \in \mathbf{C} \quad (2)$$

for the output neuron. This 1-1-1 complex-valued NN is apparently equivalent to a 2-2-2 real-valued NN (called NET 2 here) shown in Fig.1.

Proposition 1 NET 2 has no critical point based on a hierarchical structure.

(Proof) Assume a 2-1-2 real-valued NN obtained by removing the hidden neuron 1 from the NET 2 (called NET 3 here) (Fig.2). Also assume that the learning parameter of NET 3 is a critical point that implements mapping $F_1(x, y)$. It is necessary to realize any one of the following three conditions for implementing the same mapping F_1 by appending once-removed hidden neuron 1 to NET 3 again.

1) A weight vector between hidden neuron 1 appended and the two output neurons is $\mathbf{0}$.

$v = w = \mathbf{0}$ must hold in this case, but this violates the assumption $v + iw \neq \mathbf{0}$.

2) A weight vector between hidden neuron 1 appended and the two input neurons is $\mathbf{0}$.

$a = b = \mathbf{0}$ Must hold in this case, which violates the assumption $a + ib \neq \mathbf{0}$.

3) For weight vector \mathbf{w}_1 between hidden neuron 1 appended and the two input neurons, and the weight vector \mathbf{w}_2 between hidden neuron 2 and the two input neurons, $\mathbf{w}_1 = \mathbf{w}_2$ or $\mathbf{w}_1 = -\mathbf{w}_2$.

$a = b = \mathbf{0}$ Must hold in this case, but this violates the assumption $a + ib \neq \mathbf{0}$.

Therefore, mapping F_1 cannot be implemented by NET 3

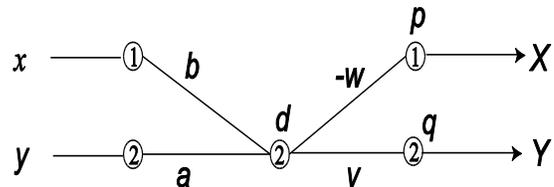


Fig. 2. NET 3, a 2-1-2 real-valued NN obtained by removing hidden neuron 1 from NET 2.

with the original hidden neuron 1 appended and having the weight structure of NET 2.

The description above illustrates a case in which hidden neuron 1 is removed, but removal of the hidden neuron 2 engenders the same conclusion. Consequently, NET 2 has no critical point based on a hierarchical structure. (QED)

See appendix A for the practical implementation process for 2-2-2 real-valued NN having no critical points based on a hierarchical structure.

B. Construction of 4-4-4 real-valued NN

A 4-4-4 real-valued NN having no critical point based on a hierarchical structure is constructed from a 1-1-1 quaternionic NN. The quaternionic NN is an extension of the classical real-valued neural network to quaternions, of which the weights, threshold values, input and output signals are all quaternions, where a quaternion is a four-dimensional number invented by W. R. Hamilton in 1843.

Consider a 1-1-1 quaternionic NN (called NET 4 here). Let the weight between the input neuron and hidden neuron be $A = a + ib + jc + kd \in \mathbf{Q}$, and the weight between a hidden neuron and an output neuron be $B = \alpha + i\beta + j\gamma + k\delta \in \mathbf{Q}$, where \mathbf{Q} represents a set of quaternions. We assume that $A \neq \mathbf{0}$ and $B \neq \mathbf{0}$. Let $C = p + iq + jr + ks \in \mathbf{Q}$ denote the threshold of the hidden neuron, $D = \mu + iv + j\rho + k\sigma \in \mathbf{Q}$ represent the threshold of the output neuron, $I = v + iw + jx + ky \in \mathbf{Q}$ be the input signal, and $O = V + iW + jX + kY \in \mathbf{Q}$ be the output signal. We can use the activation functions defined by the following equations:

$$f_o(u) = \tanh(u_1) + i \tanh(u_2) + j \tanh(u_3) + k \tanh(u_4), \quad (3)$$

$$u = u_1 + iu_2 + ju_3 + ku_4 \in \mathbf{Q} \quad (4)$$

For the hidden neuron, and

$$g_o(u) = u, \quad u \in \mathbf{Q} \quad (5)$$

For the output neuron. Because a quaternion is non-commutative for multiplication, the computational result varies with the multiplication sequence of an input value and weight: $IA \neq AI$. Accordingly, quaternion neurons of two kinds exist: a normal quaternary neuron (computing AI) and

an inverse quaternary neuron (computing IA). This paper specifically addresses a quaternionic NN that comprises only inverse quaternary neurons as an example.

In fact, NET 4 is apparently equivalent to a 4-4-4 real-valued NN (called NET 5 here) shown in Fig.3.

Proposition 2 NET 5 has no critical point based on a hierarchical structure.

(Proof) Assume a 4-3-4 real-valued NN obtained by removing the hidden neuron 1 from the NET 5 (called NET 6 here). Also assume that the learning parameter of the NET 6 is a critical point that implements mapping $F_2(v, w, x, y)$. It is necessary to realize any one of the following three conditions for implementation of the same mapping F_2 by appending once-removed hidden neuron 1 to the NET 6 again.

1) A weight vector between hidden neuron 1 appended and the four output neurons is $\mathbf{0}$.

$B = 0$ must hold in this case, but this violates the assumption $B \neq 0$.

2) A weight vector between hidden neuron 1 appended and the four input neurons is $\mathbf{0}$.

$A = 0$ must hold in this case, but this violates the assumption of $A \neq 0$.

3) Letting \mathbf{w}_j denote the weight vector between hidden neuron j and the four input neurons for any $1 \leq j \leq 4$ where the hidden neuron 1 is the appended one, then there exist some $2 \leq j \leq 4$ such that $\mathbf{w}_1 = \mathbf{w}_j$ or $\mathbf{w}_1 = -\mathbf{w}_j$

In this case, $A = 0$ must hold, which violates the assumption of $A \neq 0$.

Therefore, mapping F_2 cannot be implemented by the NET 6 with the original hidden neuron 1 appended and having the weight structure of the NET 5.

The description above presents a case in which hidden neuron 1 is removed, but removal of the hidden neuron j engenders the same conclusion ($2 \leq j \leq 4$). Consequently, NET 5 has no critical point based on a hierarchical structure. (QED)

See appendix B for the practical implementation process used for the 4-4-4 real-valued NN having no critical points based on a hierarchical structure.

C. Construction of 2-2-2 complex-valued NN

A 2-2-2 complex-valued NN having no critical point based on a hierarchical structure is constructed from a 1-1-1 quaternionic NN.

Next we consider a 1-1-1 quaternionic NN (NET 4) defined in Section II-B, assuming that

$$a + ib \neq \pm i(c + id), \tag{6}$$

$$C = 0. \tag{7}$$

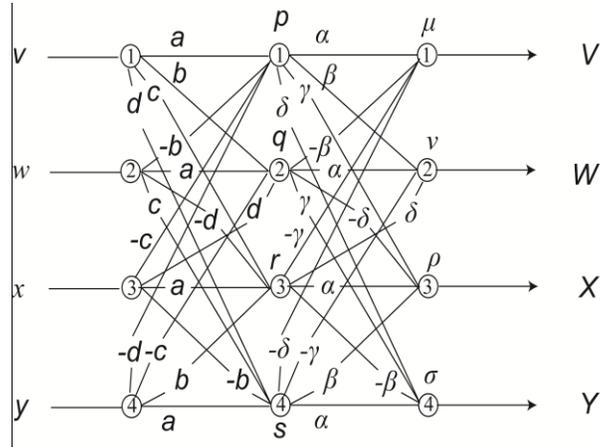


Fig. 3. NET 5, a 4-4-4 real-valued NN equivalent to the NET 4. Restriction applies between weights. $v, w, x, y, a, b, c, d, \alpha, \beta, \gamma, \delta, p, q, r, s, \mu, \nu, \rho, \sigma, V, W, X, Y$ are all real numbers.

We designate this 1-1-1 quaternionic NN as NET 7. Equation (6) has the meaning described below. The weight A between the input neuron and the hidden neuron of the NET 7 can be written using Cayley-Dickson notation as follows.

$$A = a + ib + jc + kd = x_1 + x_2j \tag{8}$$

Where $x_1 \square a + ib \in \mathbf{C}$ and $x_2 \square c + id \in \mathbf{C}$. Equation (6) can be rewritten as $x_1 \neq \pm ix_2$ from (8). Therefore, if we regard x_1 and x_2 respectively as two vectors, x_1 and x_2 do not intersect orthogonally. Furthermore, if $x_1 = \mp ix_2$, then $A = x_1 \pm x_2j = x_1(1 \pm k)$ holds. That is, the weight A has information related only to x_1 . Consequently, (6) means the exclusion of such a special case.

In addition, (7) means that the threshold value of the hidden neuron is 0, which is necessary for application of the condition for complex-valued NN being reducible.

Cayley-Dickson notation reveals that the NET 7 is equivalent to a 2-2-2 complex-valued NN (called NET 8 here) shown in Fig. 4, where $v' = v + iw, x' = x + iy, a' = a + ib, c' = c + id, \alpha' = \alpha + i\beta, \gamma' = \gamma + i\delta,$

$$\mu' = \mu + iv, \rho' = \rho + i\sigma, V' = V + iW, X' = X + iY .$$

The activation functions are given as (1) and (2).

Proposition 3 NET 8 has no critical point based on a hierarchical structure (as a complex-valued NN).

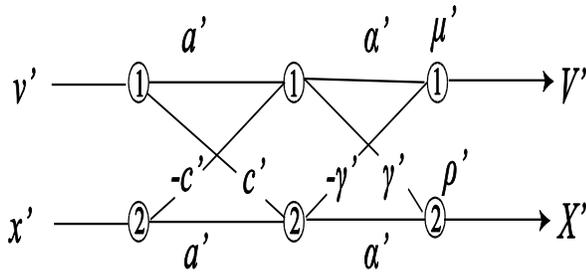


Fig. 4. NET 8, a 2-2-2 complex-valued NN equivalent to a 1-1-1 quaternionic NN (NET 7). Restriction applies between weights. $v', x', a', c', a', \gamma', \mu', \rho', V', X'$ are all complex numbers. The threshold values of the hidden neurons are all omitted because they are 0.

(Proof) Assume a 2-1-2 complex-valued NN obtained by removing the hidden neuron 1 from NET 8 (called NET 9 here). Also assume that the learning parameter of the NET 9 is a critical point that implements complex mapping $G(v', x')$. It is necessary to realize any one of the following three conditions for implementing the same complex mapping G by appending once-removed hidden neuron 1 to the NET 9 again.

1) A weight vector between hidden neuron 1 appended and the two output neurons is $\mathbf{0}$.

$\alpha' = \gamma' = 0$ Must hold in this case, but this violates the assumption $B = \alpha' + \gamma'j \neq 0$.

2) A weight vector between hidden neuron 1 appended and the two input neurons is $\mathbf{0}$.

$a' = c' = 0$ must hold in this case, but this violates the assumption $A = a' + c'j \neq 0$.

3) Let ϕ_1 be the net input to the hidden neuron 1 appended, and ϕ_2 the net input to the hidden neuron 2. Then, ϕ_1 and ϕ_2 are rotation-equivalent: $\phi_1 \equiv \phi_2$ or $\phi_1 \equiv -\phi_2$ or $\phi_1 \equiv i\phi_2$ or $\phi_1 \equiv -i\phi_2$.

Because $\phi_1(v', x') = v'a' - x'c'$ and $\phi_2(v', x') = v'c' + x'a'$,

a) Case of $\phi_1 \equiv \pm\phi_2$

$a' = c' = 0$ must hold in this case, but this violates the assumption $A = a' + c'j \neq 0$.

b) Case of $\phi_1 \equiv \pm i\phi_2$

$a' = \pm ic'$ Must hold in this case, but this violates the assumption $a + ib \neq \pm i(c + id)$ ((6)).

Therefore, the complex mapping G cannot be implemented by the NET 9 with the original hidden neuron 1 appended and having the weight structure of NET 8.

The description above presents case in which hidden neuron 1 is removed, but removal of the hidden neuron 2 engenders the same conclusion. Consequently, NET 8 has no

critical point based on a hierarchical structure (as a complex-valued NN). (QED)

This paper assumes the threshold value of the hidden neuron of a 1-1-1 quaternionic NN to be 0 ((7)). This threshold value is necessary to apply the 'three conditions for a complex-valued NN to be reducible' as described in the proof of Proposition 3. As a result, all threshold values of the hidden neuron of the obtained 2-2-2 complex-valued NN are 0. Considering a 1-1-1 quaternionic NN with possibly non-zero threshold value of a hidden neuron might yield a 2-2-2 complex-valued NN with the possibly non-zero threshold value of a hidden neuron. For a three-layer complex-valued NN with a possibly non-zero threshold value of a hidden neuron to be reducible, exceptional reducibility is necessary in addition to the three conditions presented above [2].

See the appendix C for the practical implementation process of a 2-2-2 complex-valued NN having no critical points based on a hierarchical structure.

III. DISCUSSION

Fukumizu and Amari proved that a critical point of the three-layered real-valued NN with $H-1$ hidden neurons always gives many critical points of the three-layered real-valued NN with H hidden neurons. These critical points can be local minima or saddle points.

Local minima cause *plateaus*, which have a strong negative influence on learning. Recently, it was proven that most of the local minima that Fukumizu et al. discovered are resolved by extending the real-valued NN to complex numbers; most of the critical points attributable to the hierarchical structure of the complex-valued NN are saddle points, which is a prominent property of the complex-valued NN [1]. That is, there exist many critical points based on a hierarchical structure both in the real-valued NN and the complex-valued NN.

Such critical points can be local minima or saddle points in the real-valued NN, although most critical points of the complex-valued NN are saddle points. However, in both cases, critical points do exist in the networks. As described in this paper, an attempt is made to remove critical points themselves from NNs based on a hierarchical structure.

IV. CONCLUSION

This paper presented a proposal for an implementation process of a NN having no critical point based on a hierarchical structure. Results demonstrate that real-valued and complex-valued NNs having no critical point based on a hierarchical structure can be constructed by decomposing a high-dimensional NN into equivalent real-valued or complex-valued NNs. Concretely, the following three cases are shown:

(a) A 2-2-2 real-valued NN is constructed from a 1-1-1 complex-valued NN.

(b) A 4-4-4 real-valued NN is constructed from a 1-1-1 quaternionic NN.

(c) A 2-2-2 complex-valued NN is constructed from a 1-1-1 quaternionic NN. Those NNs described above do not suffer

from a negative effect by singular points during learning comparatively because they have no critical point based on a hierarchical structure.

The author expects to address the following issues in future studies.

1) Although quaternionic NN that comprise only inverse quaternary neurons are used for this study, the case with normal quaternary neurons shall be considered.

2) General complex-valued NNs with possibly non-zero threshold values of a hidden neuron shall be analyzed, which requires consideration of exceptional reducibility [2].

3) A 2^s -dimensional Clifford NN having no critical point based on a hierarchical structure shall be produced by decomposing a general 2^n -dimensional Clifford NN [3] into equivalent Clifford NNs of 2^s dimensions ($s < n$).

V. APPENDICES

A. Implementation of the 2-2-2 real-valued NN

The practical implementation of the 2-2-2 real-valued NN having no critical points based on a hierarchical structure is described below.

1) Consider NET 1 (1-1-1 complex-valued NN) defined in Section II-A.

2) Create NET 2 shown in Fig.1 by decomposing NET 1 where the complex numbers are decomposed into two real numbers. That is, the complex number $a + ib \in \mathbf{C}$ representing the complex-valued weight between the input neuron and the hidden neuron is decomposed into the two real numbers $a \in \mathbf{R}$ and $b \in \mathbf{R}$. The complex number $v + iw \in \mathbf{C}$ representing the complex-valued weight between the hidden neuron and the output neuron is decomposed into the two real numbers $v \in \mathbf{R}$ and $w \in \mathbf{R}$. The complex number $c + id \in \mathbf{C}$ representing the complex-valued threshold of the hidden neuron is decomposed into the two real numbers $c \in \mathbf{R}$ and $d \in \mathbf{R}$. The complex number $p + iq \in \mathbf{C}$ representing the complex-valued threshold of the output neuron is decomposed into $p \in \mathbf{R}$ and $q \in \mathbf{R}$.

3) The activation functions of NET 2 are as follows:

$$f_R(u) = \tanh(u), u \in \mathbf{R} \quad (9)$$

for the hidden neurons, and

$$g_R(u) = u, u \in \mathbf{R} \quad (10)$$

for the output neurons. The following conditions are imposed on NET 2 for the assumption that $a + ib \neq 0$ and $v + iw \neq 0$ for NET 1: ($a \neq 0$ or $b \neq 0$) and ($v \neq 0$ or $w \neq 0$).

B. Implementation of the 4-4-4 real-valued NN

Practical implementation of the 4-4-4 real-valued NN having no critical points based on a hierarchical structure is the following.

1) Consider NET 4 (1-1-1 quaternionic NN) defined in Section II-B.

2) Create NET 5 shown in Fig.3 by decomposing NET 4 where the quaternions are decomposed into four real numbers. That is, quaternion $A = a + ib + jc + kd \in \mathbf{Q}$ representing the quaternionic weight between the input neuron and the hidden neuron is decomposed into the four real numbers $a \in \mathbf{R}$, $b \in \mathbf{R}$, $c \in \mathbf{R}$, and $d \in \mathbf{R}$. The quaternion $B = \alpha + i\beta + j\gamma + k\delta \in \mathbf{Q}$ representing the quaternionic weight between a hidden neuron and an output neuron is decomposed into the four real numbers $\alpha \in \mathbf{R}$, $\beta \in \mathbf{R}$, $\gamma \in \mathbf{R}$, and $\delta \in \mathbf{R}$. The quaternion $C = p + iq + jr + ks \in \mathbf{Q}$ representing the quaternionic threshold of the hidden neuron is decomposed into four real numbers $p \in \mathbf{R}$, $q \in \mathbf{R}$, $r \in \mathbf{R}$, and $s \in \mathbf{R}$. The quaternion $D = \mu + iv + j\rho + k\sigma \in \mathbf{Q}$ representing the quaternionic threshold of the output neuron is decomposed into four real numbers $\mu \in \mathbf{R}$, $v \in \mathbf{R}$, $\rho \in \mathbf{R}$, and $\sigma \in \mathbf{R}$. The activation functions of NET 5 are as follows:

$$f_R(u) = \tanh(u), u \in \mathbf{R} \quad (11)$$

for the hidden neurons, and

$$g_R(u) = u, u \in \mathbf{R} \quad (12)$$

for the output neurons.

The following conditions are imposed on NET 5 for the assumption that $A = a + ib + jc + kd \neq 0$ and $B = \alpha + i\beta + j\gamma + k\delta \neq 0$ for NET 4: ($a \neq 0$ or $b \neq 0$ or $c \neq 0$ or $d \neq 0$) and ($\alpha \neq 0$ or $\beta \neq 0$ or $\gamma \neq 0$ or $\delta \neq 0$).

C. Implementation of the 2-2-2 complex-valued NN

The practical implementation of the 2-2-2 complex-valued NN having no critical points based on a hierarchical structure is the following.

1) Consider NET 7 (1-1-1 quaternionic NN) defined in Section II-C.

Create NET 8 shown in Fig. 4 by decomposing NET 7 where the quaternions are decomposed into the two complex numbers. That is, the quaternion $A = a + ib + jc + kd \in \mathbf{Q}$ representing the quaternionic weight between the input neuron and the hidden neuron is decomposed into the two complex numbers $a' = a + ib \in \mathbf{C}$ and $c' = c + id \in \mathbf{C}$. The quaternion $B = \alpha + i\beta + j\gamma + k\delta \in \mathbf{Q}$ representing the quaternionic weight between a hidden neuron and an output neuron is decomposed into the two complex numbers $\alpha' = \alpha + i\beta \in \mathbf{C}$ and $\gamma' = \gamma + i\delta \in \mathbf{C}$.

The quaternion $D = \mu + iv + j\rho + k\sigma \in \mathbf{Q}$ representing the quaternionic threshold of the output neuron is decomposed into two complex numbers $\mu' = \mu + iv \in \mathbf{C}$ and $\rho' = \rho + i\sigma \in \mathbf{C}$. The activation functions of NET 8 are the following: (1) for the hidden neurons and (2) for the output neurons.

The following conditions are imposed on NET 8 for the assumption that $a + ib \neq \pm i(c + id)$ ((6)) and $C = 0$ ((7)) for NET 7: $a' \neq \pm ic'$, and the thresholds of the hidden neurons are all equal to zero.

ACKNOWLEDGMENT

The author would like to give special thanks to the anonymous reviewers for valuable comments.

REFERENCES

- [1] T. Nitta., "Local minima in hierarchical structures of complex-valued neural networks," *Neural Networks*, vol. 43, pp.1-7, 2013
- [2] M. Kobayashi, "Exceptional reducibility of complex-valued neural networks," *IEEE Trans. Neural Networks*, vol.21, no.7, pp.1060-1072, 2010.
- [3] E. Hitzer., T. Nitta, and Y. Kuroe, "Applications of clifford's geometric algebra," *Advances in Applied Clifford Algebras*, vol.23, no.2, pp.377-404, DOI: 10.1007/s00006-013-0378-4, 2013.